Biological Aspects of Deciphering and Engineering Regulatory Networks

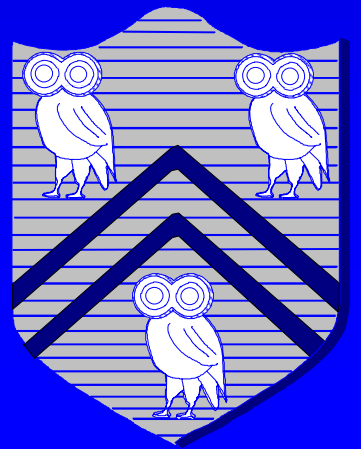
George N. Bennett

Department of Biochemistry and Cell Biology

Rice University
Houston, Texas

Metabolic Engineering Workshop
February 2007

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.



Hans V. Westerhoff and Boris N. Kholodenko: Metabolic Engineering in the Post Genomic Era

- Consider the cell as a production organism rather than as a production line.
- The organism may largely do away with much of the engineering by invoking its homeostatic control mechanisms
- Metabolic engineering should be directed at optimizing both the production flux and the functioning of the organism itself

Challenges in engineering networks of known genes & regulators

- Obtaining high levels of expression of desired genes
- Obtaining appropriate balance in levels of expression of pathway genes
- Consider regulation at enzyme level and feedback connections
- Consider the alternation of global cell physiology from the manipulation

Obtaining high levels of expression of desired genes

- Active promoters
- mRNA stability & translational efficiency
- Functionally active protein

Obtaining appropriate balance in levels of expression of pathway genes

- Library of promoters of varying strength
- Combinations of promoters & terminators, RBS for multicistronic mRNAs
- RNA regulators (antisense, RNA binding proteins)
- Plasmid vs chromosomal location

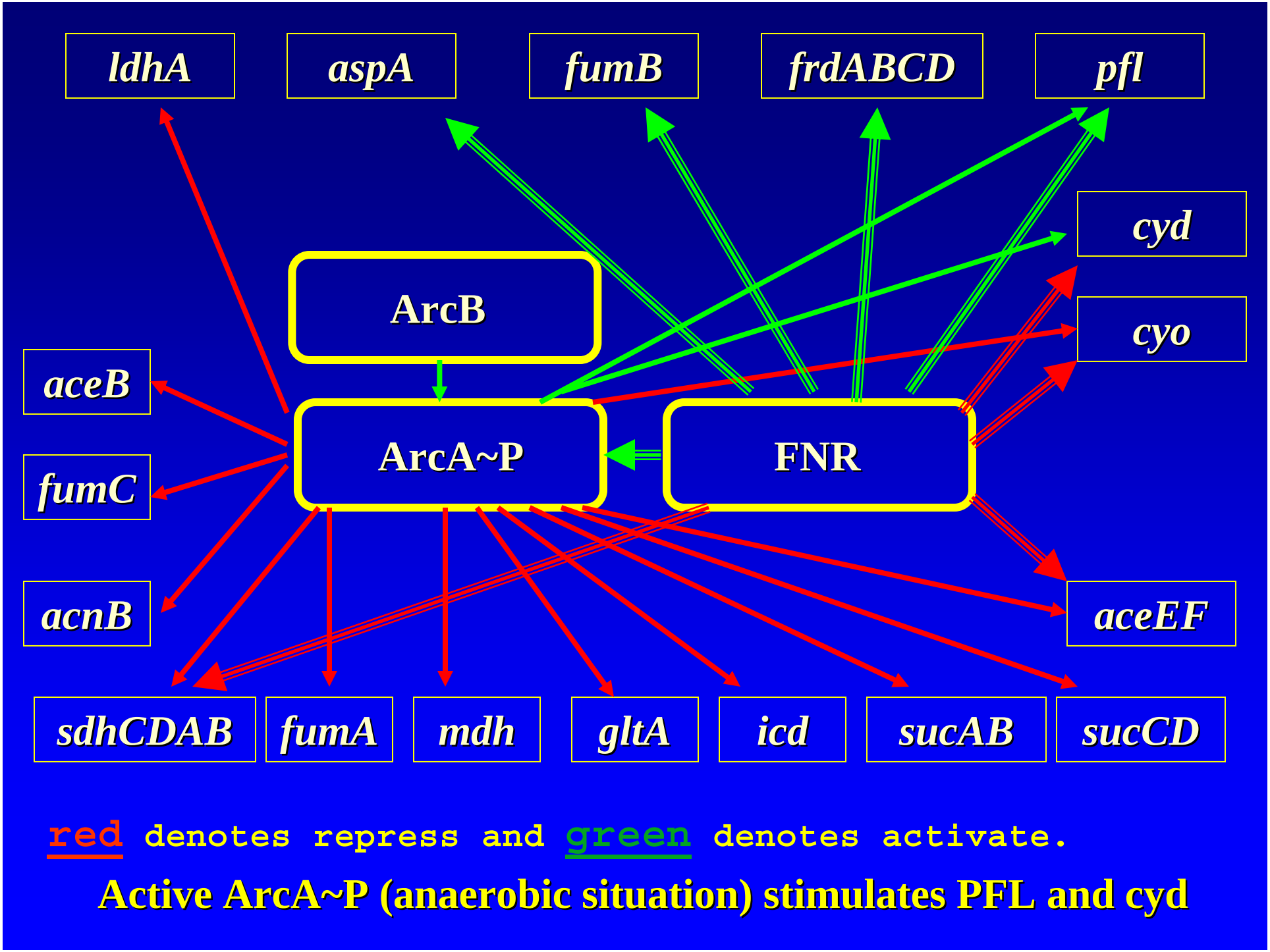
Taking into account the regulation at enzyme level and feedback connections

- **Interconnections among regulators and small molecules**
- **Picking the best enzyme from various sources or evolve such attributes**

We may know a great deal about individual components but the overall metabolic consequences of various alterations under defined conditions may not be so easily predicted

- Need detailed experiments to define ***which regulatory interactions are most important in a situation***
- Functional activity of a protein depends on more than gene expression

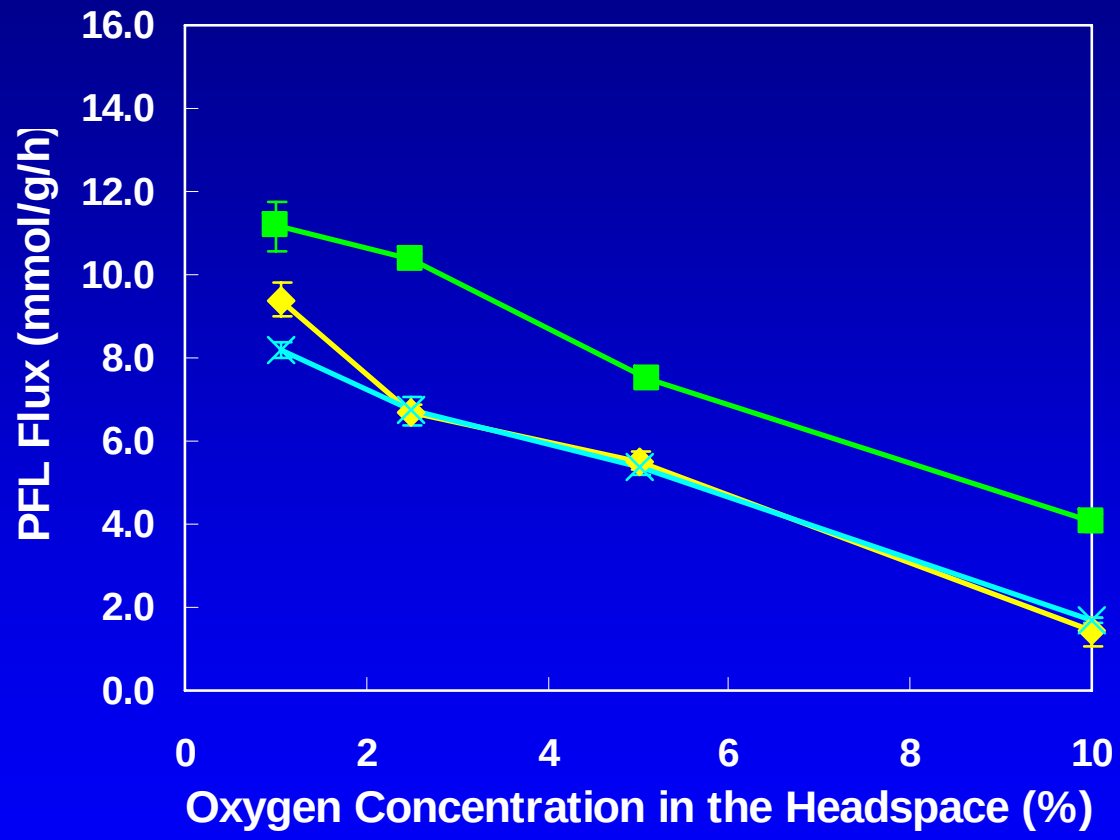
A quick example from aerobic/anaerobic E. coli



red denotes repress and green denotes activate.

Active ArcA~P (anaerobic situation) stimulates PFL and cyd

PFL flux higher in arcA mutant



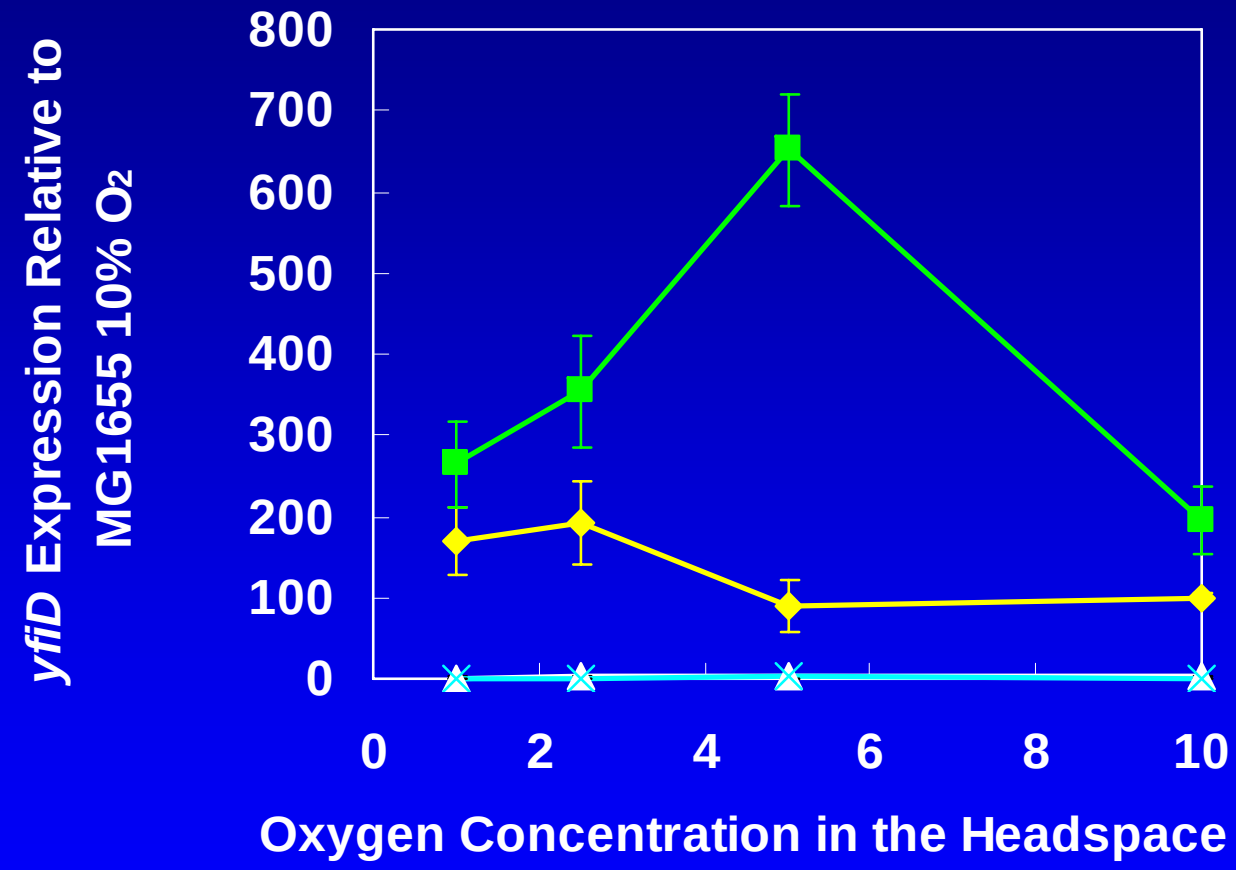
diamond
MG1655

■ MG1655 [Δ*arcA*]

× MG1655 [Δ*fnr*]

Sagit Shalel Levanon

Why are the PFL fluxes higher in the arcA mutant? (Since Arc~P activates pfl expression): PFL is reactivated by YfiD



diamond MG1655

■ MG1655 [Δ*arcA*]

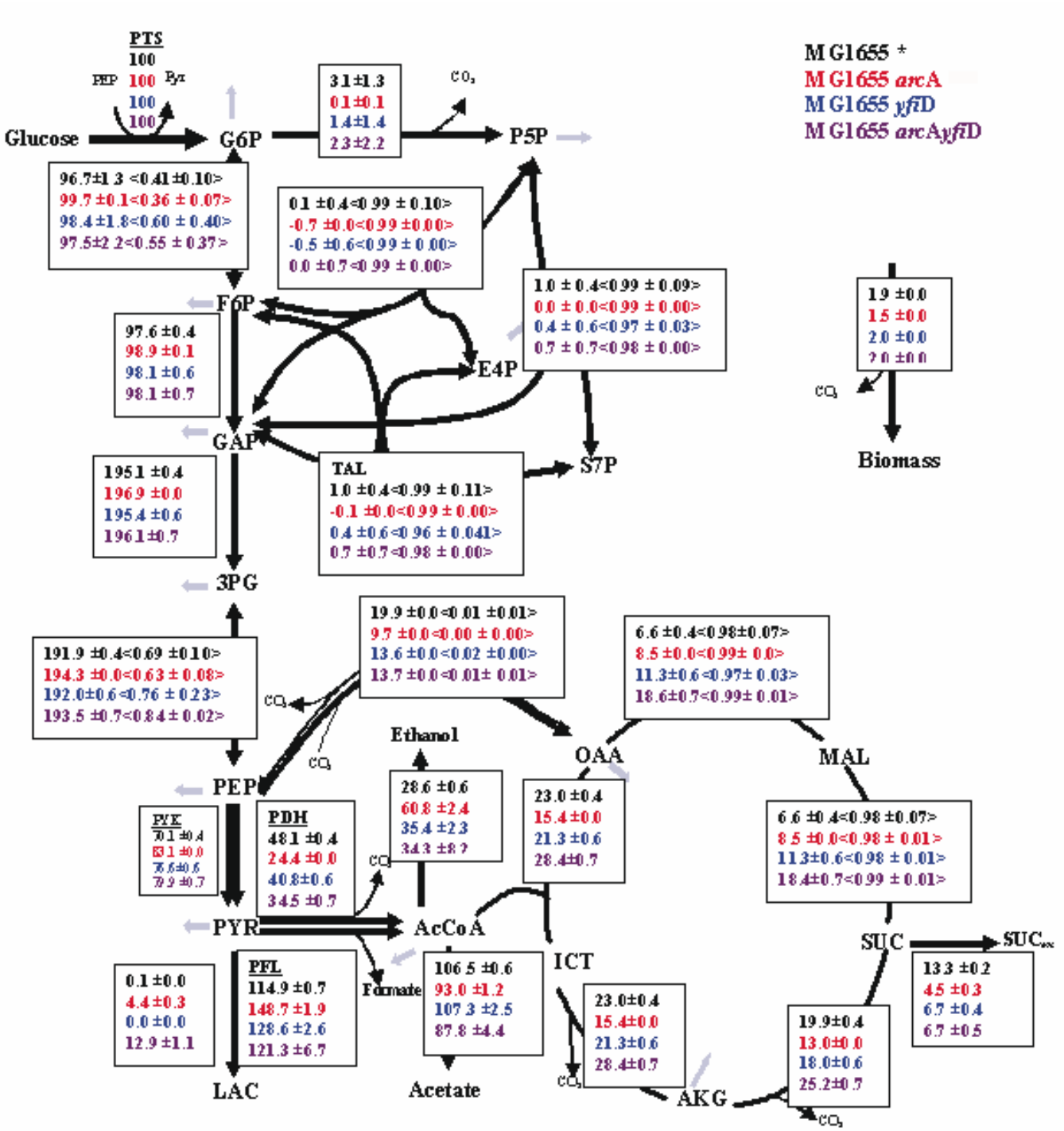
× MG1655 [Δ*fnr*]

▲ MG1655 [Δ*arcA* Δ*fnr*]

C-13 label flux analysis at 2.5% oxygen

YfiD contributes 18% of PFL flux in arcA mutant

PDH flux is ~25% of flux from pyruvate to acetyl-CoA in all strains



Overall series of actions (microaerobic)

arcA^{-} leads to lower cyd

Low cyd can not handle all NADH
and the level of reduced NADH rises

Higher NADH leads to expression of yfiD

Higher YfiD reactivates Pfl leading to higher flux through Pfl in arcA mutant in microaerobic conditions

Perspective

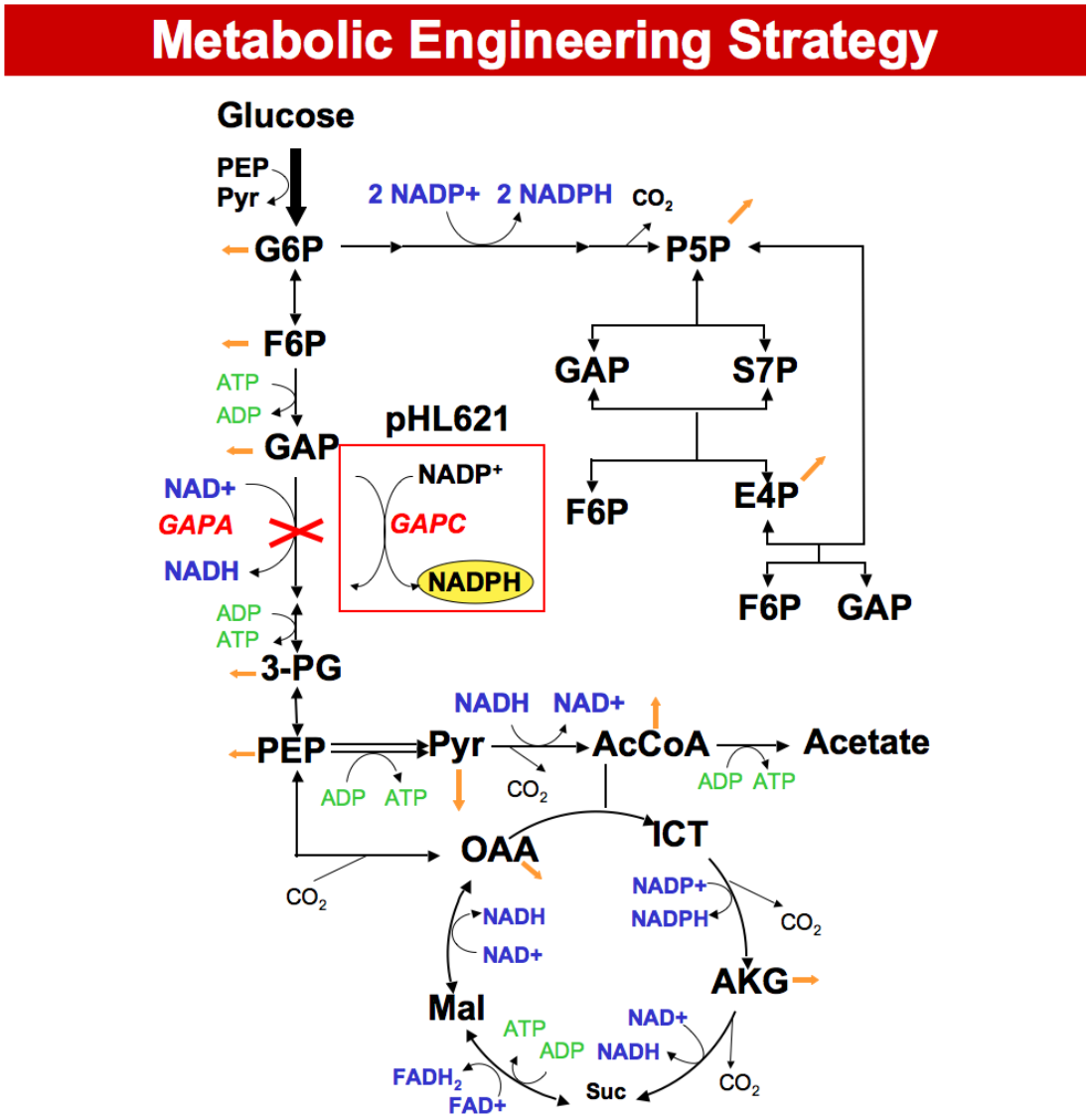
- Need a variety of experiments to see what are the important regulatory events in the network under a particular circumstance
- Methods to analyze a number of parameters in defined sample

Sometimes you can take advantage of the alternation of global cell physiology from a manipulation

- Alter levels of certain cofactors, sensors, transporters

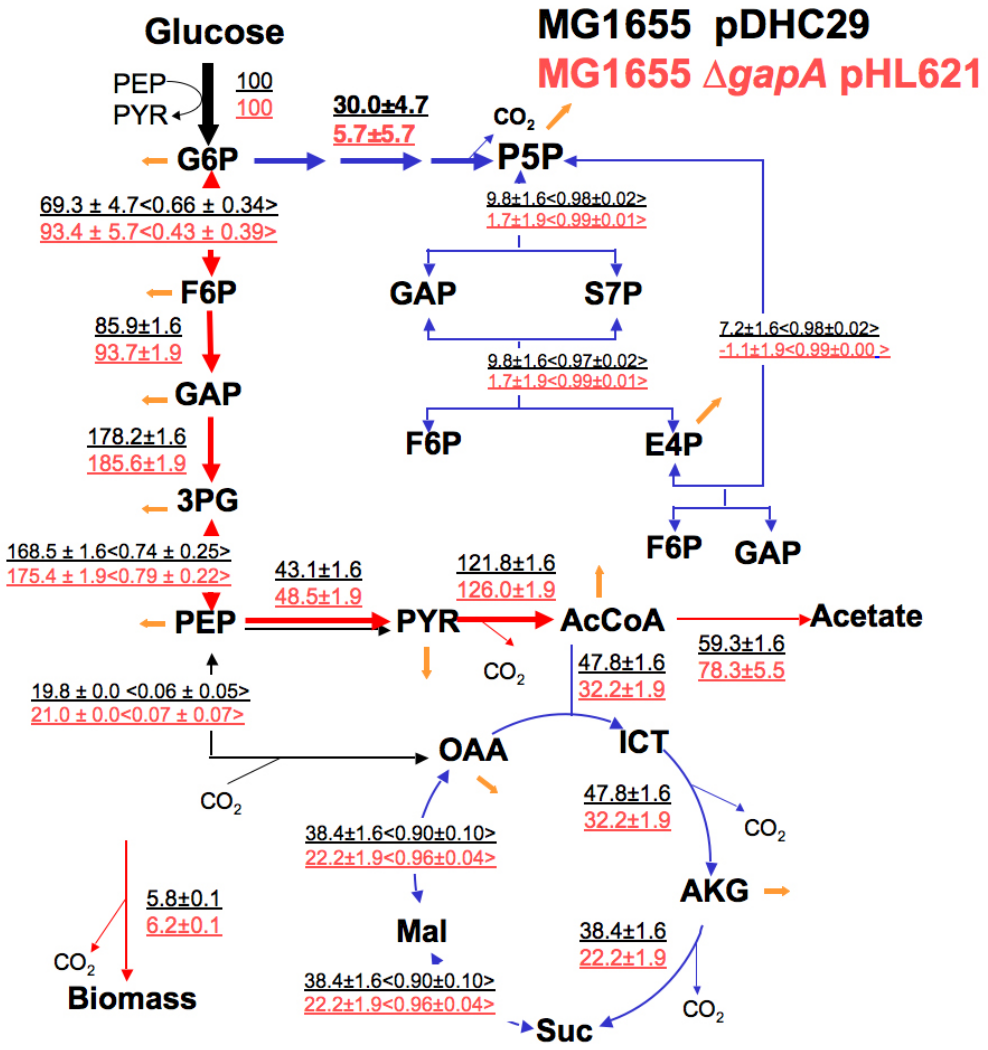
Redox reactions involving NADPH

- Engineer cells to produce more NADPH



Pathway measurements

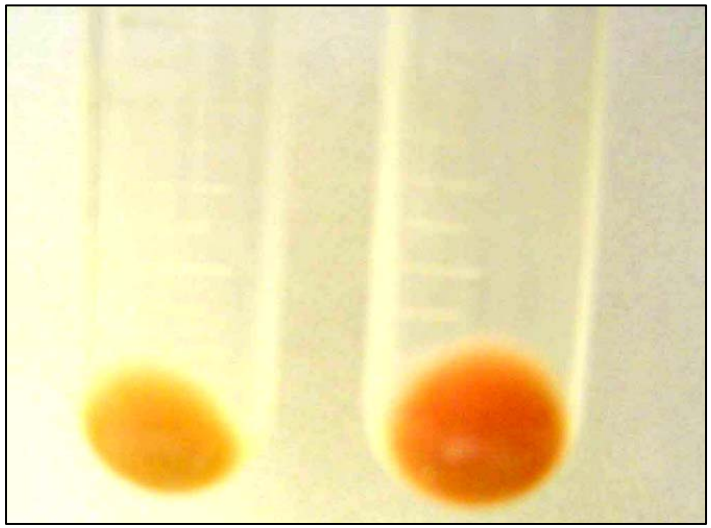
- The pentose pathway that usually supplies most of cells NADPH is much reduced in the modified stain
- Measurements made by GC-MS of C-13 labelled amino acids in steady state cultures



If there is higher NADPH availability can we use this in another pathway?

Lycopene Production Result (24 h)

WT Mutant



Challenges in engineering networks containing unknown genes & regulators

- **Identifying which genes to try to modify (define network)**
  - **Mathematical models**
  - **Gene expression microarray experiments or proteomic measurements**
  - **Tracing phosphorylation networks and connections**
  - **Similarities of protein-protein interaction networks to those of other species**

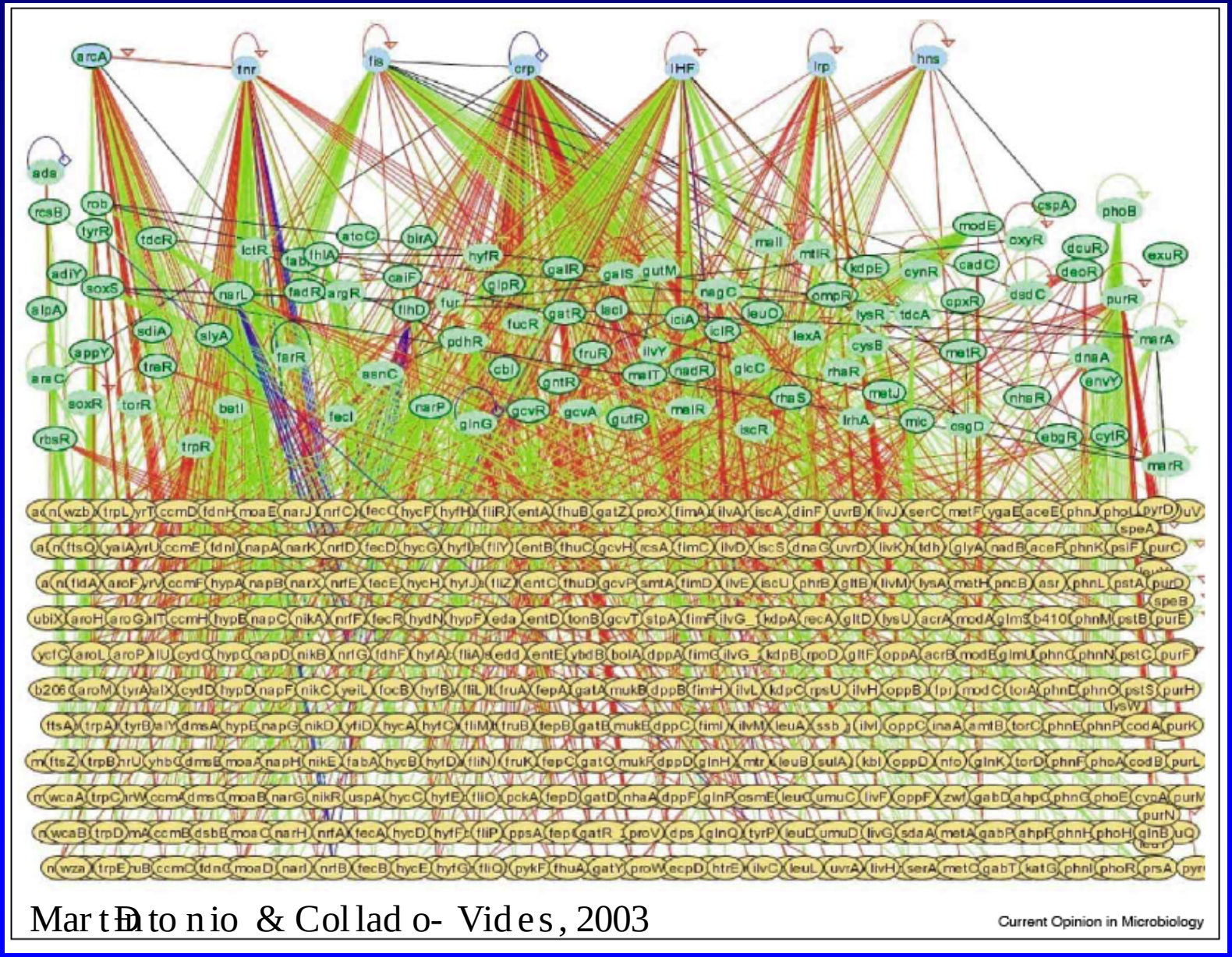
Genetic exploration

- **Individual & multiple gene knockouts or overexpression**
- **Adaptation to higher production and genome analysis**
- **Employing novel regulatory factors to affect pathway genes**
- **Use of high throughput screens, enrichments or selections**

Considerations

- **How to identify and control genes useful for a specific metabolic engineering goal?**
- **One strategy- use high-throughput means to gather data, merge in model-computational and analytical approach**
- **Another strategy- use random approach genetic based-can find things you don’t already know about**

Partial E. coli regulatory network



Martínez-Antonio & Collado-Vides, 2003

Current Opinion in Microbiology

Many global regulators affect a large number of genes

Table 1

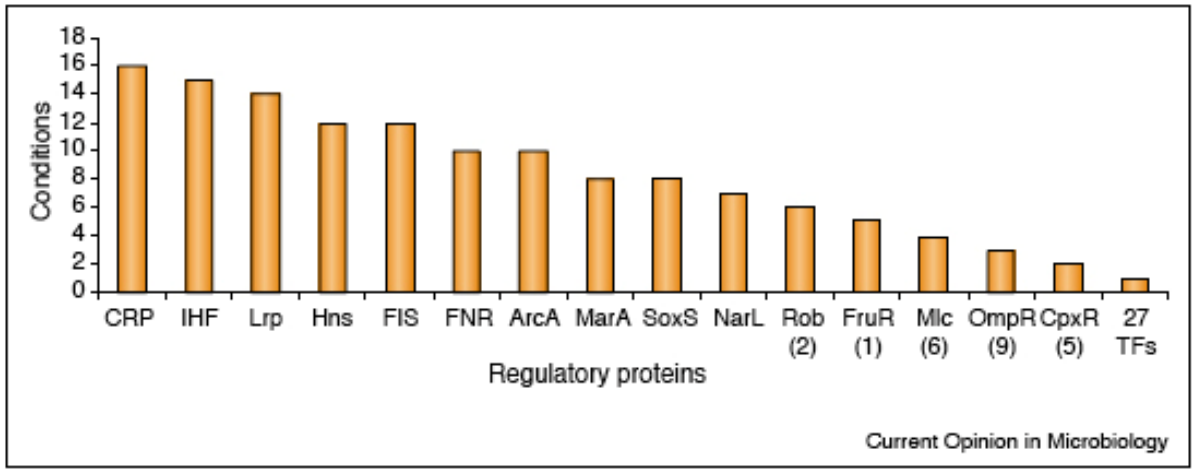
Summary of transcriptional interactions of major TFs, in the transcriptional regulatory network of *E. coli*.

Transcription factor	Genes regulated*	Co-regulators†	TFs regulated‡	Sigma factors§	Functional classes of genes regulated#	Family (members)¶
CRP	197	47	22	$\sigma^{70,38,32,24}$	48	CRP (2)
IHF	101	28	9	$\sigma^{70,54,38}$	26	HI-HNS (2)
FNR	111	20	5	$\sigma^{70,54,38}$	22	CRP (2)
FIS	76	15	4	$\sigma^{70,38,32}$	20	EBP (14)
ArcA	63	18	2	$\sigma^{70,38}$	17	OmpR (14)
Lrp	53	14	3	$\sigma^{70,38}$	15	AsnC (3)
Hns	26	14	5	$\sigma^{70,38,32}$	17	Histone-like (1)
NarL¥	65	10	1	$\sigma^{70,38,}$	14	LuxR/UhpA (17)
OmpR	10	9	3	$\sigma^{70,38}$	5	OmpR (14)
Fur¥	26	8	2	$\sigma^{70,19}$	9	Fur (2)
PhoB	26	1	3	σ^{70}	9	OmpR (14)
CpxR	9	2	1	$\sigma^{70,38,24}$	5	OmpR (14)
SoxRS	9	10	3	$\sigma^{70,38}$	10	AraC/XylS (24)
Mlc¥	5	3	1	$\sigma^{0,32}$	3	NagC/XylR (7)
CspA¥	2	2	1	σ^{70}	2	Cold (9)
Rob**	7	8	2	$\sigma^{70,38}$	6	AraC/XylS (27)
PurR**	28	7	1	σ^{70}	10	GalR/LacI (13)

*Total number of genes regulated directly. †Number of different TFs with which at least a gene or TU is jointly co-regulated. ‡Number of regulated genes that codify for TFs. §List of σ factors of the regulated promoters. #Number of functional classes of the gene products regulated [44]. ¶TF family and in parenthesis the number of members of the family. In addition to the seven global TFs considered here there are TFs suggested by ¥Babu and Teichmann, 2003, [42••] and **Shen-Orr *et al.*, 2002, [50••].

Martõntonio & Collado-Vides, 2003

Some regulators respond to many environmental growth conditions



Global environment growth conditions in which TFs are regulating. To see the detailed list of conditions see RegulonDB page: http://www.cifn.unam.mx/Computational_Genomics/regulondb/SupMat/conditions. Numbers in brackets indicate how many additional TFs participate in the same number of conditions.

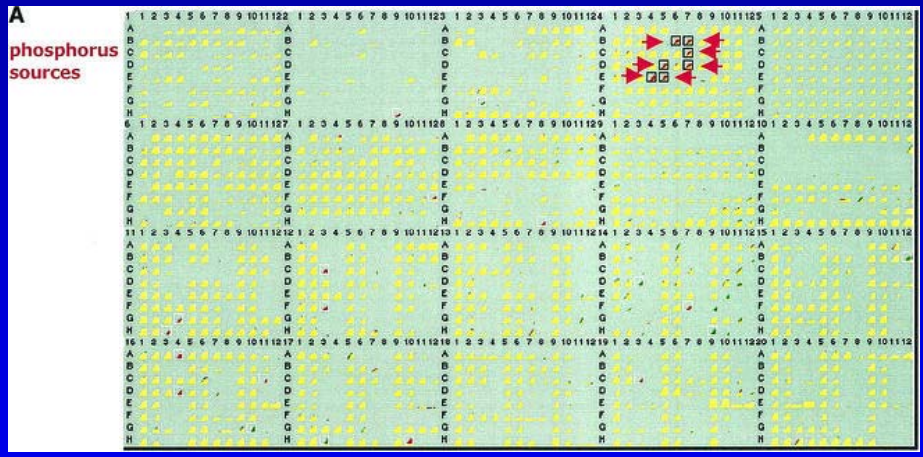
Martõntonio & Collado-Vides, 2003

Identifying important regulators in your network

- Individual specific regulators for operon
- Global regulators
- 2-component regulators
- Small RNA regulators

Test phenotypes of many regulatory genes under many different conditions-database.

- If a regulator has effect under conditions are important to production conditions; can manipulate those regulators and look for effects on the desired system?



•Zhou et al JBact 185, 4956, 2003 screened many mutants in hundreds of conditions

General Approaches to identifying functions of regulators

- What proteins bind to what gene regulatory regions in the chromosome?
- Address by variations of ChIP-chip
- What regulatory proteins interact with each other?
- Address with two-hybrid system variations or crosslinking and chemical analysis, protein interaction networks

Correlations to determine choices of genes to modify

- Two-component regulators often increased in level of their gene expression with condition where they exert their activity

Example

- **In Clostridium:**
- **Which proteins are involved in initial kinase reaction that starts solvent formation and sporulation**

If there were kinases that might phosphorylate Spo0A in *Cac*… how could we look for them?

Genome analysis: Compare genome characteristics of phosphorelay kinases with those of Bacillus subtilis

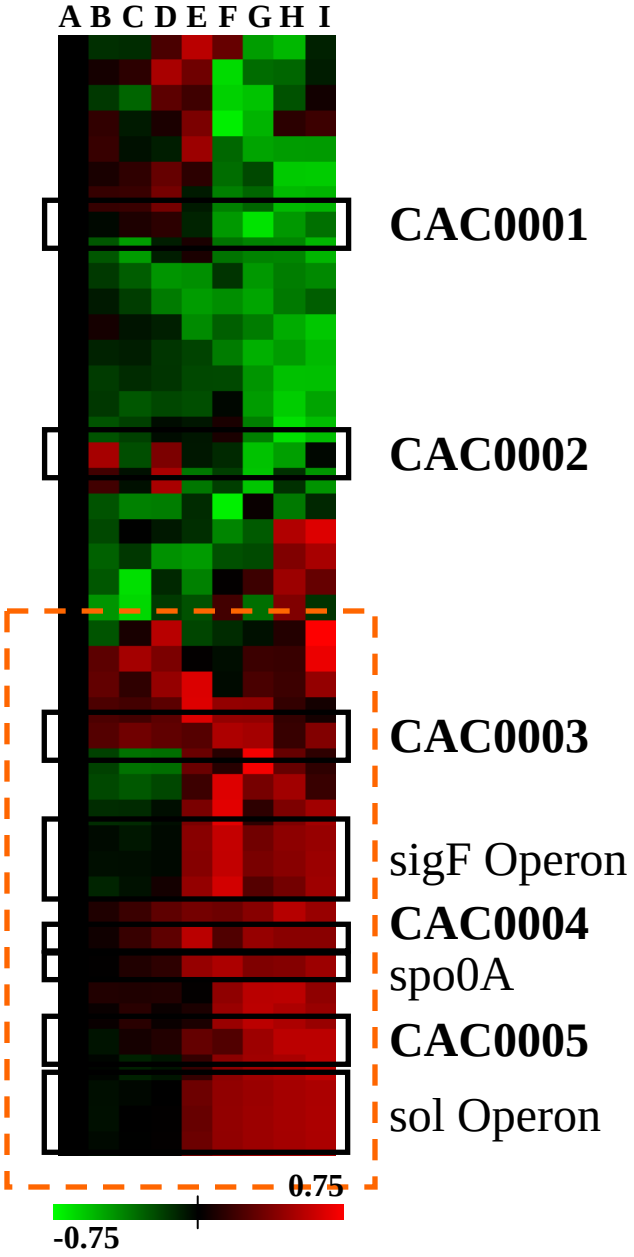
***Cac* has 36 His Kinases, 5 of them:**

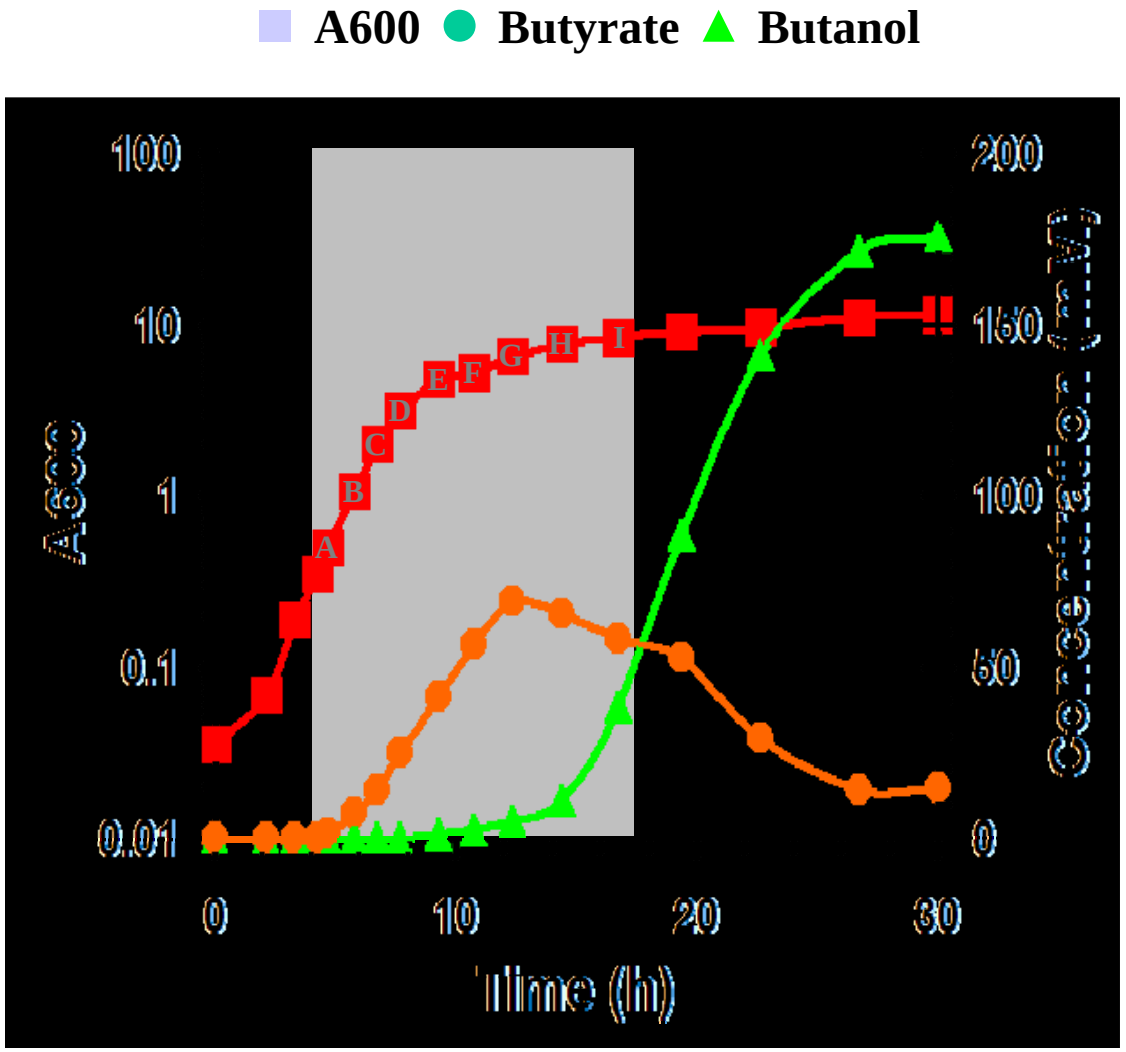
LIKELY CANDIDATES?

- **Experimental analysis:**
- **Suitable temporal expression patterns?**
- **Structural similarity?**

Papoutsakis 2005

All Cac His Kinase Expression Profiles





Orphan His Kinase

Array data after normalization, **standardization** and hierarchical clustering. Timepoint A was used as reference.

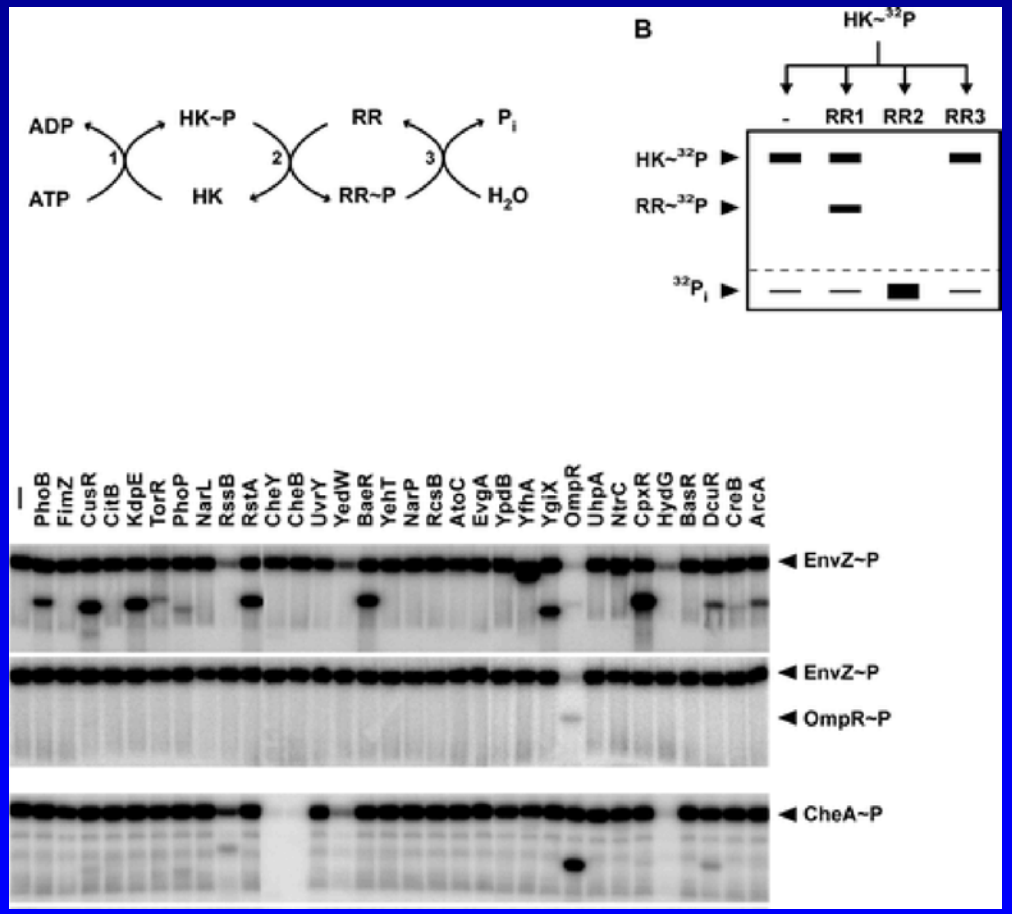
Papoutsakis 2005

Another strategy is more biochemical

- Use action of proteins in vitro to follow regulatory network

Tracing phosphorylation networks

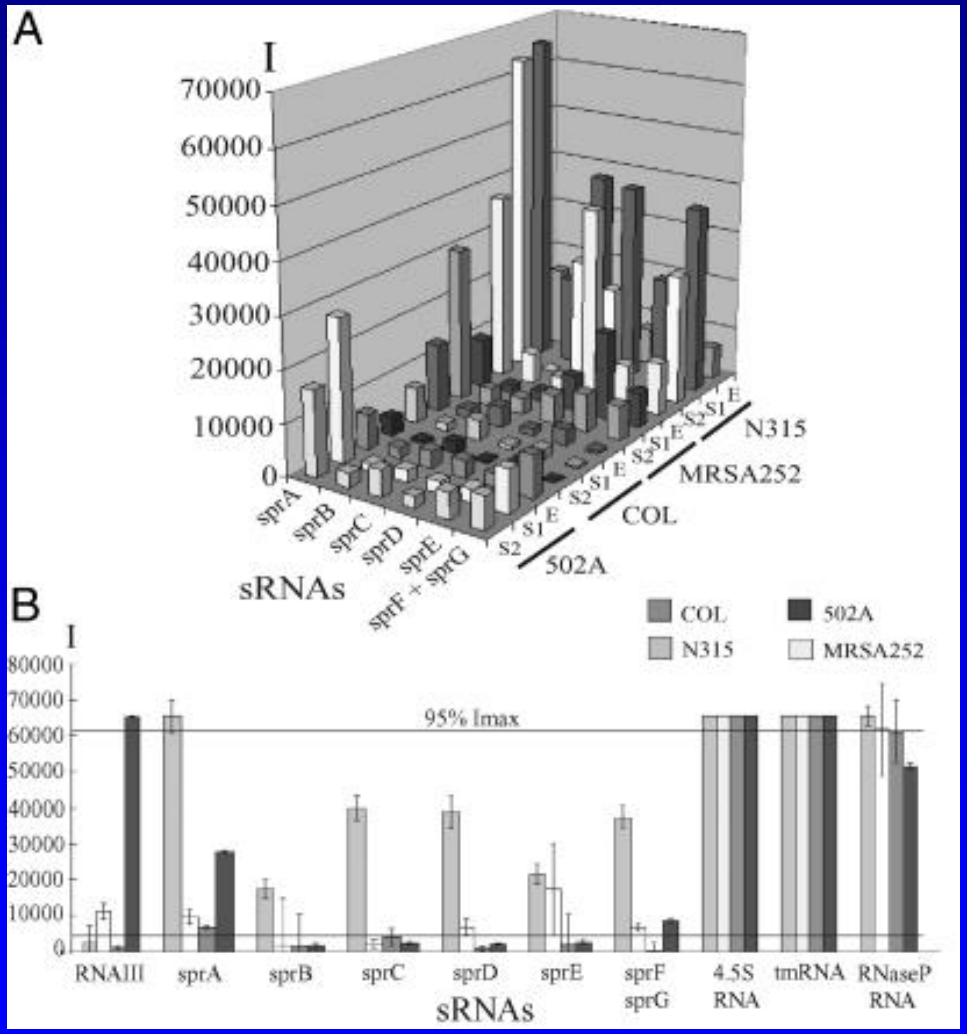
- A regulator is autophosphorylated and then tested to see what other proteins it can transfer the labeled phosphate to in vitro



•Skerker et al Plos Biology 3 (10) e334

Action of small RNAs in regulation

- Identify possible RNAs by bioinformatics
- Do microarray with these intergenic regions
- Look at expression patterns of the RNAs
- Make overexpression and deletion mutants
- Explore these mutants by microarray to identify genes under their control



Pichon & Felden PNAS 2005

Take advantage of the complexity of regulatory network

let genetics help find useful changes to make

just want improved production?

- Find out mechanism later

- Random or artificial evolution approach
- Selection or screen then analyze

Seeking effects of gene knockouts, gene overexpression, or regulatory sequence insertions

- Large scale library of variants
- Screen by optical sorting system based on metabolite or key protein to find +s

- Grow under stress of desired production condition- look for those tagged mutants found in higher abundance or greatly diminished in population

Genome sequence approach

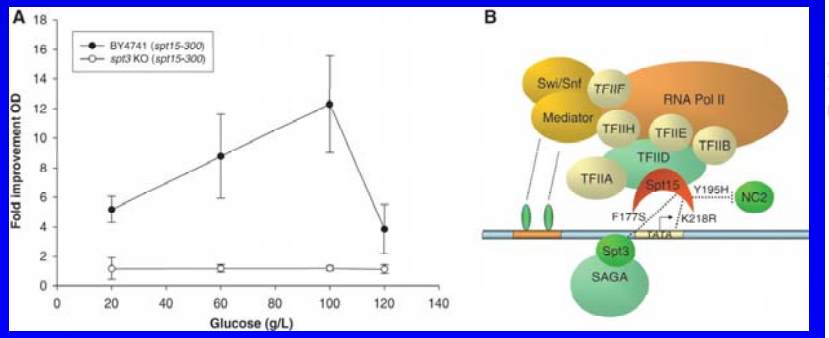
- Adaptation of cells to higher growth-production state-analyze genome of selected variant to see what has changed
- Used in industry to characterize production strain vs original parent

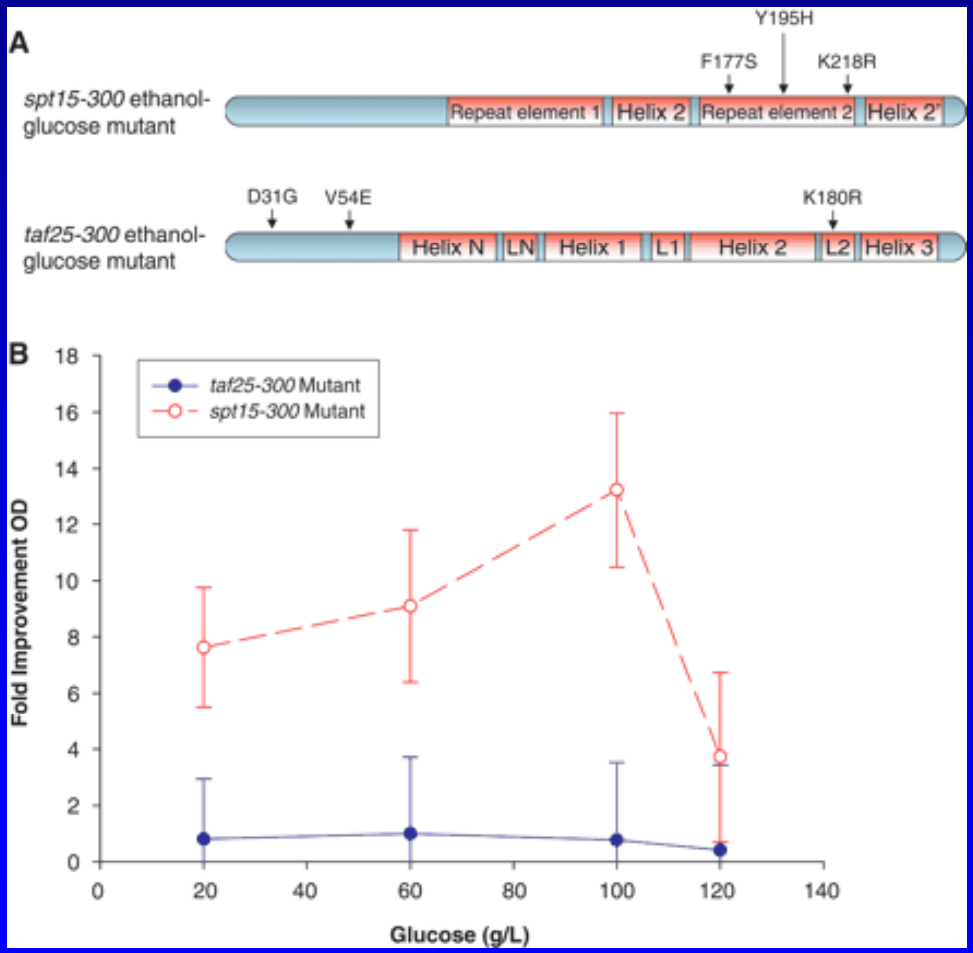
Examining the effect of novel transcriptional regulators

- Introduce mutations in known regulators
- Artificial Zn finger protein variants
- Find those that generate desired phenotype
- Then can perhaps identify and manipulate the affected genes and see how the change produced its effect

Mutations in known regulators

- **Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production**
- **Mutate TATA binding transcription factor**



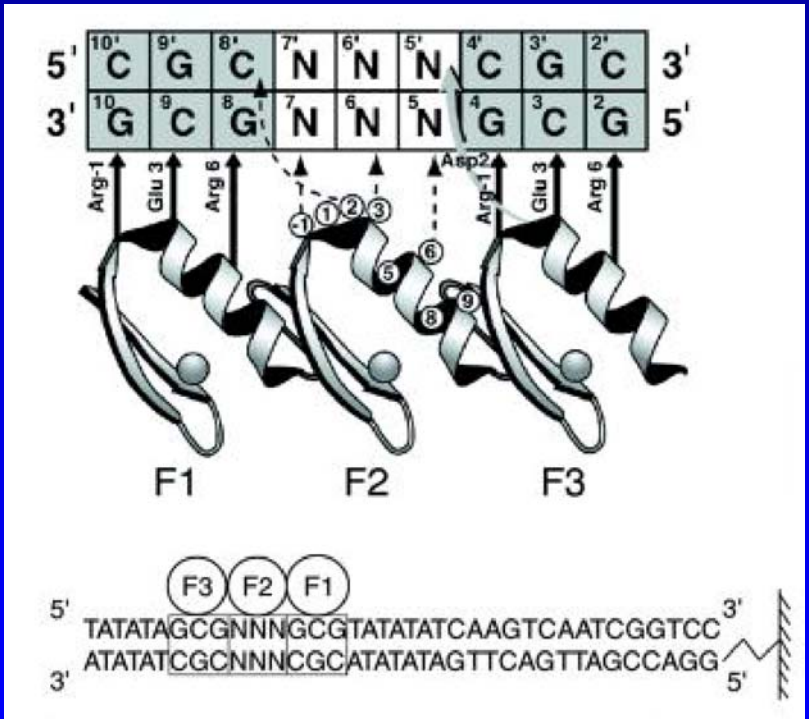


Alper et al Science 314, 1565, 2006

Zn fingers

- Directly select **artificial zinc finger** proteins from a zinc finger protein library- can join to other protein

Bae & Kim Mol Cell 376, 2006



Bulyk et al PNAS 7158, 2001

Acknowledgments

Collaborators:

Ka-Yiu San
Steve Cox
E. T. Papoutsakis

Ailen M. Sanchez
Jiangfeng Zhu

Henry Lin
Sagit Shalel Levanon

Our work was supported by grants from the National Science Foundation and USDA

(BES-0222691, BES-0000303, BES-0420840, and 2002-35505-11638)